ECE531 Screencast 8.6: Neyman-Pearson Hypothesis Testing

D. Richard Brown III

Worcester Polytechnic Institute

Examples of real-world hypothesis testing problems

- To approve a new flu test, the FDA requires the test to have a **false positive** rate of no worse than 10% and a detection rate of at least 75%.
- After major bicycling races, many riders are tested for the presence of performance enhancing drugs. The false positive rate of these tests must be less than $x\%$ and the detection rate must be at least $y\%$.
- False positives in radar systems: incoming airplane is detected as an enemy airplane when it is actually friendly. These false positives must occur with rate less than $x\%$, and the detection rate must be maximized.

In many hypothesis testing problems, there is a **fundamental asymmetry** between the consequences of

- "false positive" (decide $\mathcal{H}_1$ when the true state is x_0) and
- "miss / false negative" (decide $\mathcal{H}_0$ when the true state is x_1).

Neyman-Pearson Terminology

Neyman-Pearson hypothesis testing is always binary (simple or composite).

$$\begin{aligned} \mathcal{H}_0 &: \text{“null” hypothesis or “signal absent”} \\ \mathcal{H}_1 &: \text{“alternative” hypothesis or “signal present”} \end{aligned}$$

Common terminology for simple binary hypothesis testing:

- A “type I error” is when you decide $\mathcal{H}_1$ when the state is x_0. Also called a “false alarm” or “**false positive**”.

$$R_0(D) = \mathrm{Prob}(\text{decide } \mathcal{H}_1 | \text{state is } x_0) = P_{\mathsf{fp}}(D)$$

- A “type II error” is when you decide $\mathcal{H}_0$ when the state is x_1. Also called a “miss” or “**false negative**”.

$$R_1(D) = \mathrm{Prob}(\text{decide } \mathcal{H}_0 | \text{state is } x_1) = P_{\mathsf{fn}}(D)$$

- The “power” of a test is the probability of correctly deciding $\mathcal{H}_1$ when the state is x_1 or, in other words,

$$\text{power} = \mathrm{Prob}(\text{true positive}) = 1 - \mathrm{Prob}(\text{false negative}) = P_{\mathsf{D}}(D)$$

  The power of the test is also the probability of detecting the signal is present.

The Neyman-Pearson Criterion

Definition

The Neyman-Pearson criterion decision rule is given as

$$\begin{aligned} \rho^{\mathsf{NP}} &= \arg\max_{\rho \in \mathcal{D}} P_{\mathsf{D}}(\rho) \\ \text{subject to } P_{\mathsf{fp}}(\rho) &\leq \alpha \end{aligned}$$

where $\alpha \in [0, 1]$ is called the "significance level" of the test.

This is a "constrained optimization" problem.

Note that maximizing P_D is equivalent to minimizing the conditional risk $R_1(D)$.

Neyman-Pearson Hypothesis Testing Example

Coin flipping problem with a probability of heads of either $q_0 = 0.5$ or $q_1 = 0.8$. We observe three flips of the coin and count the number of heads. We can form our conditional probability matrix

$$P = \begin{bmatrix} 0.125 & 0.008 \\ 0.375 & 0.096 \\ 0.375 & 0.384 \\ 0.125 & 0.512 \end{bmatrix} \text{ where } P_{\ell j} = \mathrm{Prob}(\text{observe } \ell \text{ heads} | \text{state is } x_j).$$

Suppose we need a test with a significance level of $\alpha = 0.125$.

- What is the N-P decision rule in this case?
- What is the probability of correct detection if we use this N-P decision rule?

What happens if we relax the significance level to $\alpha = 0.5$?

NP Decision Rule (part 1 of 2)

Main idea: Sort the likelihood ratios $L_\ell = \frac{P_{\ell,1}}{P_{\ell,0}}$ in descending order. The order of L's with the same value doesn't matter.

So in our three-coin flip problem, we have

$$L_{\text{sorted}} = [L_3, L_2, L_1, L_0]^\top = [4.1626, 1.0240, 0.2560, 0.0640]^\top$$

The Neyman-Pearson decision rule for simple binary hypothesis testing with discrete observations is then:

$$\rho^{\mathsf{NP}}(y) = \begin{cases} 1 & \text{if } L_\ell > v \\ \gamma & \text{if } L_\ell = v \\ 0 & \text{if } L_\ell < v \end{cases}$$

Need to specify the likelihood threshold v and randomization γ...

NP Decision Rule (part 2 of 2)

The decision threshold $v \geq 0$ is the minimum value such that

$$P_{\mathsf{fp}} = \sum_{\ell : L_\ell > v} P_{\ell,0} \leq \alpha.$$

Note the strict inequality.

Once you have v, you have a decision rule δ^v that satisfies the false positive probability constraint. If this constraint is satisfied with equality, then $\gamma = 0$ and you are done. Otherwise, you need to determine the randomization coefficient $\gamma \in [0,1]$. The false positive probability is

$$P_{\mathsf{fp}} \quad = \quad (1-\gamma) P_{\mathsf{fp}}(\delta^v) + \gamma P_{\mathsf{fp}}(\delta^{v-\epsilon})$$

Setting this equal to α and solving for γ yields

$$\gamma = \frac{\alpha - P_{\mathsf{fp}}(\delta^v)}{P_{\mathsf{fp}}(\delta^{v-\epsilon}) - P_{\mathsf{fp}}(\delta^v)} = \frac{\alpha - \sum_{\ell : L_\ell > v} P_{\ell,0}}{\sum_{\ell : L_\ell = v} P_{\ell,0}}$$